DOCUMENT RESUME

ED 164 559 TM 007 527

AUTHOR Secarea, Rebecca V.; Olsen, Larry K.
TITLE An Evaluation Instrument for Appraising the Health Knowledge of Seventh Grade Students Participating in a Special School Health Education Project.
PUB DATE 13 Oct 74
NOTE 12p.; Paper presented at the Annual Meeting of the American School Health Association (48th, New York, New York, October 13, 1974)
AVAILABLE FROM Larry K. Olsen, Chair, Health Science, Arizona State University, Tempe, Arizona 85281 (for a copy of the test)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS *Achievement Tests; Educational Assessment; Grade 7; *Health Education; Junior High Schools; Multiple Choice Tests; *Test Construction; Test Reliability; Test Validity
IDENTIFIERS School Health Education Curriculum Project

ABSTRACT

A multiple-choice test was developed to assess cognitive gains resulting from participation in the School Health Education Curriculum Project. The test was intended for seventh grade students, and was based on curricular objectives in the cognitive domain: knowledge, comprehension, application, and evaluation. Educators who were familiar with the project curriculum reviewed the items, and trial tests were administered to 159 students. After the items were reviewed for acceptable discrimination, difficulty, distractor functioning, skewness, and reliability, a 47-item revised test resulted. One conclusion of this study was that the sentence completion method of developing distractor responses for multiple choice items provided a practical means of retaining the thoughts and vernacular of students for whom the test was intended. Additional testing was recommended to acquire additional data on reliability, validity, and norms. It was also suggested that more items be developed in the areas of analysis, synthesis, and evaluation. (GDC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

AN EVALUATION INSTRUMENT FOR APPRAISING THE HEALTH KNOWLEDGE OF SEVENTH GRADE STUDENTS PARTICIPATING IN A SPECIAL SCHOOL HEALTH EDUCATION PROJECT

Rebecca V. Secarea, M.S.
Instructor, East Leyden High School
Franklin Park, Illinois

Larry K. Olsen, Dr. P.H.
Assistant Professor, Health Education
University of Illinois
Department of Health & Safety Education
Champaign, Illinois 61820

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Larry K. Olsen
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

ED164559

TM007527

Introduction

Accountability has become the "watchword" in educational circles today. It is important that educators be able to demonstrate that stated course objectives are being met. Further, it must be shown that courses which students must "pass" are relevant and necessary in today's changing world.

Health education is assuming a more formidable position in today's changing school curriculum. This trend is noticeable when one examines the amount of legislation either pending or which has recently been passed requiring concentrated health instruction as an integral part of the total school curriculum. In order to meet the responsibility of showing that course objectives are being met through the various health education courses, educators must be prepared to offer positive evaluation of course objectives which demonstrate that students are, in fact, gaining new information, and, have the ability to apply that knowledge. As a result, it is necessary that valid and reliable evaluation instruments be developed based upon the objectives set forth for a particular course. Too often a new course of study is developed and evaluation becomes an afterthought. Evaluation is

---

This paper was presented before the Research Council at the forty-eighth Annual Convention of the American School Health Association in New York, N.Y., October 13, 1974.

attempted with hastily developed teacher made tests or, for the sake of expediency, an instrument which was previously developed and may appear to be applicable to the course in question is utilized. This suggests that our programs are stagnating and not providing dynamic learning experiences or keeping up to date with current information in the health field. Secondly, too often these instruments have been designed for evaluation of objectives other than those in which we are interested.

Purpose

The purpose of this study was to construct a valid and reliable evaluation instrument that would assess the health knowledge of seventh grade students participating in a prototype health education curriculum.

Scope

The scope of this study was defined using the seventh grade unit of study from the School Health Education Curriculum Project. The unit consists of six major sections, each with its own focus and basic goals. The course is designed for approximately twelve weeks of instruction, five days per week for approximately one hour per day. At present this curriculum is being utilized in approximately 120 school districts throughout the United States.*

Significance

The significance of this study falls into four major categories:

1. The test instrument is developed specifically for use with the seventh grade unit of study of the School Health Education Curriculum Project.

2. The study fulfills a need for a standardized instrument to asses cognitive gains as a result of participation in the School Health Education Curriculum Project.

---

*For a discussion of this project, see Roy L. Davis, "Making Health Education Relevant and Exciting in Elementary & Junior High School," HSMA Reports, U. S. Department of Health, Education & Welfare, Vol. 88, No. 2, (Feb.), P. 99-108, 1973.

3. The study provides teachers with some new methodologies they might employ when constructing evaluation instruments for their various classes.

4. The instrument provides an objective basis for revision of the health curriculum.

5. The study provides teachers with one form of evaluation for determining teacher or program effectiveness.

Procedures

In order to develop an evaluation instrument commensurate with course objectives, it is important that we know the educational outcomes expected of students. This necessitates that course content and educational objectives be well defined and specified. Most test developers would agree that the first step in test development is a precise definition of educational objectives. The course of study was carefully reviewed and a total of twelve major content areas were identified. Specific objectives within each content area were also delineated. The twelve content areas were utilized in developing a table of specifications for the test instrument. The major function of the table of specifications is to illustrate the number of test items needed to obtain a balance between content and behavioral objectives.[1]

To augment the table of specifications, and assure that test items fell into a taxonomy of cognition, Bloom's six subdivisions of the cognitive domain were reviewed.[2] The six subdivisions consist of knowledge, comprehension, application, analysis, synthesis and evaluation. The synthesis and analysis categories were eliminated in the final analysis since none of the items developed by students tended to fall into these categories. Further, it is usually quite difficult for even the most experienced test developer to construct multiple choice items for these two categories.[3]

The remaining four subdivisions were placed on the horizontal axis of the table of specifications and the twelve content areas were placed on the vertical axis. Using this table as a guide, stems were developed which would be representative of each of the content areas and which would fall into each of the four divisions of the taxonomy.

The 120 stems thus developed were submitted to three of the six national training directors from the School Health Education Curriculum Project, four classroom teachers involved with the project and four faculty members of the Department of Health and Safety Education of the University of Illinois, Urbana - Champaign who were familiar with the project curriculum. They were asked to make any revisions or additions they felt necessary. This, combined with the table of specifications, helped ensure content validity.

Often when constructing multiple choice items, the test developer analyzes what should be covered in a test, writes a series of stems, then proceeds to develop foils or distractors with little regard for their suitability, plausibility or vernacular. Naturally, this procedure is expeditious, however one can not be sure he is developing items which are representative of student cognition about a particular area of concern. To combat this problem, the 120 stems developed for this study were randomly placed into three booklets of 40 items each. These booklets took the form of a sentence completion test. Rimland and Zwerski[4] found that this methodology is a reliable procedure for obtaining a universe of statements which could be utilized in constructing multiple choice tests.

The three sentence completion tests were then submitted to approximately 100 seventh grade students from Seattle, Washington, who had recently

completed the unit of study. The test forms were randomly assigned to students and the students were asked to complete each sentence. Upon completion of the test forms, they were returned to the researchers for analysis.

The sentence completion items were grouped by test form and each item was analyzed to ascertain the "universe" of response for that particular item. Of the 120 stems originally developed, 50 appeared to have some degree of response agreement among the trial population. These items fell into all twelve content areas and the four subdivisions of Bloom's Taxonomy. In addition, there was agreement among the jury of experts that these items were representative of both the content and objectives of the unit.

The responses of the students were compiled into multiple choice form with four foils for each stem. Except for grammatical and verb tense changes, the foils were left in the vernacular of the students. In constructing the trial instrument the recommendations of the construction experts were followed.[5]

The items in the trial test were arranged in order of difficulty progressing from lower to higher cognitive skills.[6] The purpose for doing this is to avoid confusion or discouragement for the student early in the testing process. The trial instrument was precoded to facilitate keypunching and minimize transfer errors.

Since the test was to be the beginning of a standardized instrument, three national training locations were selected to participate in the tryout phase of the study. These three were selected because they all had students who were currently involved in the project and the teachers agreed to cooperate with the study. The sites selected were, Seattle, Washington; El Cajon, California; and Charleston, West Virginia. Approximately

300 students were included in the trial population.

The test instrument was administered by the classroom teacher immediately upon completion of the unit of instruction. Test forms were returned to the researchers, given identification numbers and submitted to the Survey Research Laboratory of the University of Illinois for keypunching. Analysis of test data was accomplished using the MERMAC System Total Test and Item Analysis Package developed by the Office of Measurement and Research of the University of Illinois.[7]

Analysis of Test Data

Although the total sample size was 300, the test forms from Seattle were lost in the mail leaving a total sample of 159 students for final analysis. This was recognized as a major limitation of the study.

Distractor functioning, difficulty rating and discrimination index were considered to be major factors in the development of the revised test instrument. Acceptable limits for each of these was set as follows:

1. a discrimination index of +.20
2. a difficulty rating between .10 and .90
3. a minimum distractor functioning of three percent

Items not meeting the above standards were considered for revision. However, because a particular item did not meet all standards did not mean that the item was discarded. In analyzing the data both aggregate and geographic location were considered. The reason for this was that although on aggregate data a particular item may not meet predetermined standards, it may function quite well within a geograph c location. Further, the small sample size would tend to make one retain an item which approached the standard in order that additional testing might be done. This situation may be a direct result of population characteristics or of variations in curriculum content from one geograpic location to another. This occurred with respect to three items pertaining to sex education. Since sex education per se was not taught in the school system in Charleston, items

pertaining to this area of health education were omitted from the test instrument used there.

The possible range of test scores was zero to fifty. The obtained range was 9 to 34. Skewness of the test was calculated at +.04. Positive skewness is an indication that the test was slightly difficult since skewness is an indication of the relationship between high and low test scores. If the median falls below the mean, then high scores are having a stronger influence on the test mean than are the smaller scores. In a perfect distribution, the skewness would be zero. The skewness of most tests falls between plus or minus 3, thus the skewness of the test was well within acceptable limits.

To assess test reliability, Kruder-Richardson Formulas 20 and 21 were utilized. These formulas are a measure of the magnitude of relationship between the various items on the sample tested. One problem in utilizing these formulas is that they assume that all test items are homogenous and that all test items are measuring the same general factors of ability. The greater the variance in item difficulty, the greater both these formulas tend to underestimate test reliability. The KR20 for this test was 0.536 and the KR21 was 0.424. Simce the test items were designed to meet four areas of cognition, they could not be considered to be homogenous. In addition, since some of the items did not meet predetermined test specifications, they exerted a negative effect. As a result, the reliability, as estimated by KR20 and 21 is an underestimate of the actual test reliability and further testing is necessary to determine true test reliability. Revision of test items should help yield a reliability coefficient above .60 as recommended by Levitt for group comparison purposes.[8]

A total of twenty test items met all criteria established for the test. When data were analyzed by geographic location, an additional eleven

items met the criteria. The three items dealing with family living deleted from the West Virginia test met all criteria in the California sample. These items were retained for the revised instrument. Two items were found to be relatively easy for both sample populations with 91 and 93 percent of the sample answering each item correctly. These items were also retained for the revised instrument since they did measure important concepts and contributed to test validity. Further, items such as this act as motivational devices for students taking the test.[9]

A total of thirteen items received some form of revision because they failed to meet the minimum discrimination standard of +.20. One item was revised because it did not fall within the difficulty range of .10 to .90. Five items had distractors not meeting the standard of three percent. Of the initial fifty items, only three were discarded because they failed to meet standards in more than one of the three criterion areas. This resulted in a revised test of 47 items.

Limitations

The following were recognized as factors which limited this study:

1. Participating districts had to slightly modify the curriculum so that it fit within their local board of education guidelines. As a result, some test items were not relevant in specific geographic locations.

2. Approximately half of the original test data was lost in shipment, therefore only two of the three geographic locations were represented for analysis of the trial test for the purpose of making revisions.

3. The revised instrument has not been field tested, therefore its reliability has not been established.

Summary

The purpose of this study was to construct a valid and reliable evaluation instrument that would assess the health knowledge of seventh grade students participating in a prototype health education curriculum.

The unit of study was reviewed for content and objectives and a table of specifications for the trial instrument was developed. Multiple choice was selected as the best test method. Item stems were developed which met the requirements set forth in the table of specifications. The stems were submitted to persons familiar with the unit of instruction for comment and revision. The revised stems were randomly placed into three test booklets and submitted to three school districts participating in the curriculum.

Test booklets were returned to the researchers and a series of fifty multiple choice items using the student responses as distractors were developed. These items were submitted to a jury for comment. The items were then placed into a test booklet and distributed to the three school districts. The test was administered to students immediately upon completion of the unit of instruction. Completed test forms were returned to the researchers for analysis by the Division of Measurement and Research of the University of Illinois.

Based upon aggregate data, twenty items were found to meet all test specifications and were retained in their original form for the revised instrument. Three items were omitted from one sample but met test specifications in the other sample. Consequently, these items were retained for the revised instrument. Eleven items failed to meet test specifications when analyzing aggregate data but did meet the specifications when analyzed by geographic location. As a result, these items were retained in their original form. A total of thirteen items received some form of revision. Three items were rejected from the revised form because they failed to meet the predetermined test standards in more than one area. This resulted in a final revised form of 47 items.

Conclusions

A major conclusion drawn from this study was that the sentence completion method of developing distractors for multiple choice items provides a practical

means of retaining the thoughts and vernacular of students for whom the test is intended.

Secondly, the revised test provides seventh grade teachers with a standardized, objective instrument for evaluating cognitive changes in their students as a result of participating in the School Health Education Curriculum Project. Finally, the instrument provides educators with objective criteria upon which they can base curriculum revisions or indicate areas within the existing curriculum which need additional emphasis.

Recommendations

As a result of this study the following recommendations are offered:

1. Further testing of the revised instrument using a larger sample size is needed to establish reliability.

2. Upon completion of additional testing, norms for the test should be developed.

3. Additional effort should be made to develop items dealing with analysis, synthesis and evaluation and these items should be included in the instrument.

4. A suitable criterion variable such as the AAHPER Cooperative Health Education Test should be correlated with the revised instrument to establish criterion validity.[10]

5. The sentence completion method of developing test items should be employed by more test developers thereby helping eliminate researcher bias when developing health education tests.

References

1. Gronlund, N.E.: Stating Behavioral Objectives for Classroom Instruction. New York: The Macmillan Co., 1970, p. 41.

2. Bloom, B.S.: Taxonomy of Educational Objectives. New York: David McKay Co., 1965.

3. Solleder, M.K.: "Evaluation in the Cognitive Domain," Journal of School Health, XLII(1): 16-20, (Jan), 1972.

4. Rimland, B., and E. Zwerksi: "The Use of Open End Data as an Aid in Writing Multiple-Choice Distractors," Journal of Applied Psychology. Vol. 46 (1): 31-33, (Jan.), 1962; L. K. Olsen: "An Evaluation Instrument for Assessing the Health Related Attitudes of College Students." Journal of School Health. XLII (7): 408-411, (Sept.), 1972.

5. Board, C. and D. Whiting: "The Effect of Selected Poor Item Writing Practices on Test Difficulty, Reliability and Validity." Journal of Educational Measurement. (Fall): 225-233, 1972; R. L. Ebel: Essentials of Educational Measurement. Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1970; Educational Testing Service: "Multiple Choice Questions: A Close Look.": Measurement and Evaluation in Psychology and Education. R. Thorndike and E. Hagan (eds.). New York: John Wiley and Sons, Inc., 1969; and J. C. Nunnally: Educational Measurement and Evaluation. New York: McGraw-Hill Book Co., 1972.

6. Conrad, H. S.: "The Experimental Tryout of Test Materials." in Educational Measurement. E. F. Lindquist (ed.), Washington, D.C.: American Council on Education, 1963.

7. Office of Measurement and Research: "Test and Questionnaire Analysis Programs Written for the IBM System/360." in Mermac Manual. Urbana: University of Illinois Press, 1971.

8. Levitt, E.: Clinical Research in the Behavioral Sciences. Springfield, Illinois: Charles P. Thomas, 1961, p. 32.

9. Wesman, A. G. "Writing the Test Item," in Educational Measurement. op. cit; and Stan Henrysson: "Gathering, Analyzing and Using Data on Test Items," in Educational Measurement. ibid.

10. American Association for Health, Physical Education and Recreation: Cooperative Health Education Test. Berkeley, California: Cooperated Tests and Services, 1971.